(12)

Special Report 85-4

A COMPARISON OF DICHOTIC LISTENING TASK SCORING METHODS

G. R. Griffin and J. D. Mosko

March 1985

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
PENSACOLA FLORIDA

Approved for public release; distribution unlimited.

A COMPARISON OF DICHOTIC LISTENING TASK SCORING METHODS

G. R. Griffin and J. D. Mosko*

Naval Medical Research and Development Command
63706N M0096001.1051

Reviewed by

F. E. Guedry, Ph.D.
Senior Scientist

Approved and Released by

Captain W. M. Houk, MC USN
Commanding Officer

March 1985

Naval Aerospace Medical Research Laboratory
Naval Air Station, Pensacola, Florida 32508-5700

SUMMARY PAGE

THE PROBLEM

In a recent evaluation of two dichotic listening tasks (DLTs) as predictors of performance in Naval Aviation Undergraduate Pilot Training, results based on one of five scoring methods were presented. The purpose of this report is to provide DLT performance scores using all five scoring methods for comparative purposes to determine the scoring system most economical and efficient for automated scoring and, most sensitive to individual and mean differences.

FINDINGS AND RECOMMENDATIONS

Five scoring methods are described which vary primarily in their treatments of errors, and in their consideration of the importance of sequence effects. Five independent analyses of previously reported data were performed. Results were nearly invariant across all five methods; i.e., intercorrelations among scores across scoring techniques exceeded .90. Two of the scoring methods are recommended for purposes of standardizing future analyses of DLT performance, one because of its simplicity and ease of application and the other, because it may have greater sensitivity to differences in individual performance.

Acknowledgments

Appreciation is extended to the following individuals for their valuable comments and suggestions in the preparation of this report: LT Dennis K. McBride, MSC USNR, Gerald B. Thomas, Ph.D., and Carl E. Williams, Ph.D.

*Dr. J. D. Mosko's current address is Rome Air Development Center, Griffiss Air Force Base, New York 13441.

INTRODUCTION

Aviation personnel must attend to various arrays of simultaneous as well as sequential information inputs. The number of potential sources of information and the rapidity of information flow often require efficient switching of attention among sensory modalities and among channels within modalities. Success in allocating attention to appropriate sources may account for a significant share of variance in piloting success.

An attentional shifting dichotic listening task (DLT) developed by Gopher (1) and Gopher and Kahneman (2) proved to be a valid predictor of student performance in aviator training for the Israeli Air Force. The DLT task requires subjects to maintain attention to target and irrelevant vocal information presented to a designated ear, ignore information presented to the other ear, and subsequently, to recognize or recall the target information presented via the designated ear. The reported merits of the DLT have gained the regard of a number of investigators (4, 5, 6), many of whom are attempting to increase the efficiency of test batteries used to predict success for students entering military flight training programs in the United States.

In a recently reported U.S.Navy study (3), results indicated that a dichotic listening task similar to the one used by Gopher (1) accounted for a statistically significant portion of pass-fail variance in the Navy flight training program. The results were based, however, on only one of five scoring methods developed by the investigators at the Naval Aerospace Medical Research Laboratory (NAMRL). The primary variations among these scoring techniques relate to (a) the treatment of error scores (intrusions, omissions, etc.) and (b) the treatment of sequence effects. Table 1 provides a description of the contrasting features of the five scoring schemes. It is conceivable that the application of different scoring methods might yield different conclusions. The purpose of this study is to examine this possibility and document the correlational relationships between the various scoring methods.

TABLE 1

DLT Scoring Methods and Performance Measures[1]

Scoring Method Number	Performance Measures
1. An error-based, sequence-dependent scoring method with differentiation error type. These errors are labeled ommissions, intrusions, and other (after Gopher & Kahneman (2)). An omission error is the failure to report a relevant digit in a designated ear. An intrusion error is the reporting of a digit from the nondesignated ear. "Other" errors include the reporting of a letter or a digit when one is not presented or incorrectly reporting a digit. All responses are scored.	* Part 1 omissions, intrusions and other error * Part 2 omissions, intrusions and other errors * Part 1 number correct = 180 - Total Part 1 errors * Part 2 number correct = 144 - Total Part 2 errors * Total correct = Part 1 number correct + Part 2 number correct (nine performance measures)
2. A modified error-based sequence-dependent scoring method highly similar to scoring method 1 with only a slight variation in the treatment of intrusion errors. In this scoring method an intrusion error normally results in one additional (omission) error.	* A total of 9 performance measures similar to scoring method 1
3. A modified sequence-independent simple number correct scoring method. All responses are scored regardless of sequence or the number of responses.	* Part 1 number correct Part 2 number correct Total number correct
4. A modified sequence-independent simple number correct scoring method where only the first five responses of Part 1 and the first four responses of Part 2 of each DLT trial are considered.	* Three performance measures similar to scoring method 3
5. A strict sequence-dependent simple number correct scoring method. Once an error occurs no additional correct responses are awarded. Part 1 and Part 2 responses are scored independently.	* Three performance measures similar to scoring method 3

[1] Examples of the results of the application of each of the different scoring systems are presented at Appendix A.

METHOD

Apparatus and Procedure. The DLTs, extensively described elsewhere (3), consisted of the dichotic presentation of letter-digit strings. The subject was instructed to maintain attention to one ear while ignoring the information presented to the other ear, and subsequently, to record on an answer sheet the digits presented to the designated ear, accurately and in the sequence of occurrence. The test apparatus comprised a dual-channel tape recorder, headphones, and a paper-and-pencil answer booklet. A VOTRAX synthetic speech system was utilized to generate the auditory stimuli at sound pressure levels of approximately 75 dB/Leq (re.: 20 μ P).

Each DLT trial was divided into two parts (see Figure 1). The Part 1 task consisted of a mix of letters and digits delivered to each ear. Digits were never presented simultaneously to the two ears, and no digit was repeated in either sequence. However, simultaneous presentations were presented of identical or dissimilar letters, or a letter to one ear and a digit to the opposite ear. Part 2 of each trial consisted of the simultaneous presentation of two letters to each ear followed by a string of four successive digits. Part 1 and Part 2 of each trial were each preceeded by a "right" or "left" vocal command, signifying the designated ear. The auditory stimuli were presented at the rate of one letter or digit per 0.9 second. A single trial, including pause time, lasted approximately 27 seconds. One hundred eighty total correct responses were possible for Part 1, 144 for Part 2, over 36 trials. A diagram of a sample DLT trial is at Appendix B.

PART 1

Left Ear R 8 N S M Y 2 G B 7 F L 6 R L 5

"Right" (Vocal Channel "attend" Command)

Right Ear Y L 3 S R 4 F Z 9 X F Ø F N 1 L

PART 2

Left Ear B F 4 3 7 9

"Left" (Vocal Channel "attend" Command)

Right Ear G L 1 5 6 2

Figure 1. DLT Trial Example

Methodological departures from the original Israeli DLT were (a) the use of letter text rather than Hebrew verbs, (b) the use of "left" and "right" vocal channel "attend" commands presented stereophonically (rather than tones presented monaurally to the "attend" ear), and (c) the use of computer generated speech with simultaneous stimulus onset times, left and right ear, (rather than the tape recorded voice of a female speaker), and in addition, (d) the requirement for written, rather than oral, responses.

Preliminary research suggested that the DLT lacked sufficient difficulty. A number of attempts to increase the level of difficulty were tried (e.g., tones as channel attend commands, background "party" distracting speech, and varying loudness levels), though unsuccessfully. Finally, it was demonstrated that incorporating irrelevant background materials (digits recorded in reverse--"zero" becomes "orez") to each channel at a sound pressure level equal to that of the relevant test material significantly increased test error. The initial DLT and the DLT containing background material were designated the Clear DLT (CDLT) and the Background DLT (BDLT), respectively. The only difference between the two was that the added background material was applied to the BDLT.

Subjects. Ninety-four male Navy and Marine Corps student naval aviators (SNAs) awaiting assignment to undergraduate training volunteered to participate in the study. Seventy subjects were Marine SNAs from the Marine Aviation Training Support Group at the Naval Air Station, Pensacola, Florida and 24 were Navy SNAs from the Naval Aviation Schools Command at the Naval Air Station, Pensacola, Florida.

The CDLT sample consisted of 12 Navy and 34 Marine Corps SNAs. The BDLT comprised 12 Navy and 36 Marine Corps SNAs.

RESULTS

Each of the five scoring methods described in Table 1 was applied to the DLT scores for each subject. The correlations among the various performance parameters (number correct, omission errors, intrusion errors, etc.) based on scoring method 1 for CDLT and BDLT are presented in Tables 2 and 3, respectively. Most of the intercorrelations are quite strong with the exception of Table 3 "other errors". "Other errors," incidently, comprise the smallest percentage of the various error categories for both the CDLT and the BDLT, based on scoring system 1. "Omissions" are the largest error category (CDLT 51%, BDLT 58%) followed by "Intrusions" (CDLT 33%, BDLT 29%) and then "Other errors" (CDLT 16%, BDLT 12%). The magnitude of the correlations suggest that except for "other errors" of the BDLT the various performance scores are in general agreement.

Table 2

CDLT Intercorrelation Matrix (Scoring Method 1)

DLT Measures		1	2	3	4	5	6	7	8	9
Part 1 Correct	1	-								
Part 2 Correct	2	.922	-							
Total Correct	3	.982	.978	-						
Part 1 Omission Error	4	-.975	-.926	-.973	-					
Part 1 Intrusion Error	5	-.932	-.824	-.899	-.863	-				
Part 1 Other Error	6	-.687	-.613	-.665	.605	.550	-			
Part 2 Omission Error	7	-.887	-.981	-.950	.906	.787	.534	-		
Part 2 Intrusion Errors	8	-.857	-.966	-.927	.803	.762	.574	.951	-	
Part 2 Other Errors	9	-.743	-.688	-.731	.702	.690	.632	.571	.530	-

Table 3

BDLT Intercorrelation Matrix (Scoring Method 1)

DLT Measures		1	2	3	4	5	6	7	8	9
Part 1 Correct	1	-								
Part 2 Correct	2	.710	-							
Total Correct	3	.914	.935	-						
Part 1 Omission Error	4	-.946	-.776	-.925	-					
Part 1 Intrusion Error	5	-.883	-.581	-.780	.747	-				
Part 1 Other Error	6	-.340	-.044	-.146	.086	.287	-			
Part 2 Omission Error	7	-.713	-.965	-.922	.791	.549	.040	-		
Part 2 Intrusion Errors	8	-.626	-.953	-.866	.696	.538	.136	.881	-	
Part 2 Other Errors	9	-.263	-.205	-.251	.170	.301	.405	.26	.075	-

The primary statistical concern focuses upon the family of relationships among scores derived under the five differing scoring techniques. Because of the large number of performance measures associated with the scoring methods and, as reported above, because of the strong relationships among performance data, the scores derived under each scoring method were collapsed to provide a single total number correct score for purposes of analysis. Normative values for the scoring methods are presented at Table 4. Repeated measures analysis of variance statistical treatments indicate significant differences between scoring methods. For the CDLT [$F(4,180) = 33.94$; $p < .01$], scoring method 5 resulted in significantly more errors, while scoring methods one, two, three, and four produced similar results. For the BDLT [$F(4,188) = 178.55$; $p < .01$], methods two and four, and three and four produced similar results. All other scoring method comparisons were significantly different.

The ratio of standard deviation to mean provides one measure of the sensitivity of the various scoring techniques to individual differences in performance. Using this measure as a standard (see Table 4) scoring method 5 is more sensitive to individual variation for both the CDLT and BDLT.

More important than comparison of normative values, the high intercorrelations (all are .9+) among total correct scores (see Table 5) indicate that the rank ordering of scores is generally insensitive to the scoring technique employed. Undoubtedly, the high degree of similarity between scoring methods one and two and methods three and four contributed to the high correlation values for these measures. The high positive correlations indicate, in other words, that a hypothetical rank ordering of the performance scores for subjects ranging from lowest to highest, remains intact regardless of the scoring technique selected.

Table 4

Means, Standard Deviations and Standard Deviation/Mean Ratios of DLT Performance Scores Obtained with the Five Scoring Methods

	Mean	Standard Deviation	St.Dev./ Mean
Clear DLT (46 subjects) Scoring Method			
1	315.Ø	14.8	.Ø47
2	316.2	11.8	.Ø37
3	318.8	9.5	.Ø3Ø
4	317.9	9.7	.Ø31
5	3Ø5.3	24.Ø	.Ø79
Background DLT (48 subjects) Scoring Method			
1	277.9	28.3	.1Ø2
2	286.8	21.Ø	.Ø73
3	294.7	18.3	.Ø62
4	291.7	19.Ø	.Ø65
5	25Ø.1	36.1	.144

Table 5

Intercorrelations Among DLT Performance Scores Obtained with the Five Scoring Methods

	Scoring Method	1	2	3	4	5
Clear DLT (46 subjects)	1	-				
	2	.995	-			
	3	.983	.974	-		
	4	.994	.991	.99Ø	-	
	5	.975	.983	.952	.978	-
Background DLT (48 Subjects)	1	-				
	2	.984	-			
	3	.974	.974	-		
	4	.977	.987	.98Ø	-	
	5	.9Ø4	.945	.9ØØ	.946	-

CONCLUSIONS

Based on the high positive correlations between the various scoring methods, it must be concluded that individual variation is generally insensitive to the scoring method employed. Thus, the selection of scoring technique probably has no bearing on the interpretations or generalizations derived from application of the selected technique. This means that the conclusions provided by Griffin and Mosko (3) are confirmed.

The simplicity of scoring method 3 (automated or manual), however, makes it the more attractive technique from the standpoint of economy. This method is a sequence-independent response technique. Although the technique does not differentiate error type, it is completely objective, and it is relatively easy to score as Table 1 suggests. Scoring method 5, on the other hand, is appealing because of its apparently higher sensitivity to individual variation as measured by the ratio of its standard deviation to mean. This scoring method, unlike method 3, is sequence-dependent (i.e., only correct responses preceding initial errors in Part 1 and Part 2 are awarded). Scoring method 5 is completely objective although decidedly more difficult to score than is scoring method 3.

REFERENCES

1. Gopher, D. A selective attention test as a predictor of success in flight training. Human Factors, 1982, 24:173-183.

2. Gopher D. and Kahneman, D. Individual differences in attention and the prediction of flight criteria. Percep. and Mot. Skills, 1971, 33:1335-1342.

3. Griffin, G. R. and Mosko, J. D. Preliminary evaluation of two dichotic listening tasks as predictors of performance in Naval Aviation Undergraduate Pilot Training. NAMRL 1287, Pensacola, Florida:Naval Aerospace Medical Research Laboratory, July 1982.

4. Imhoff, D. L. and Levine, J. M. Development of a perceptual-motor and cognitive performance task battery for pilot selection. Report ARRO-3049-FR, AFHRL Contract F33615-79-C-0004, Advanced Research Resources Organization, March 1980.

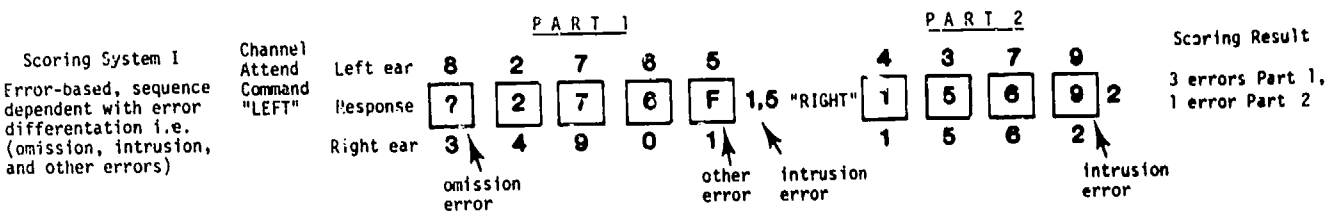
5. Myers, D. C., Schemmer, F. M. and Fleishman, E. A. Analysis of computer interactive tests for assigning helicopter pilots to different missions. Technical Report R-83-8, Army Research Institute Contract MDA 903-82-C-0036, Advanced Research Resources Organization, May 1983.

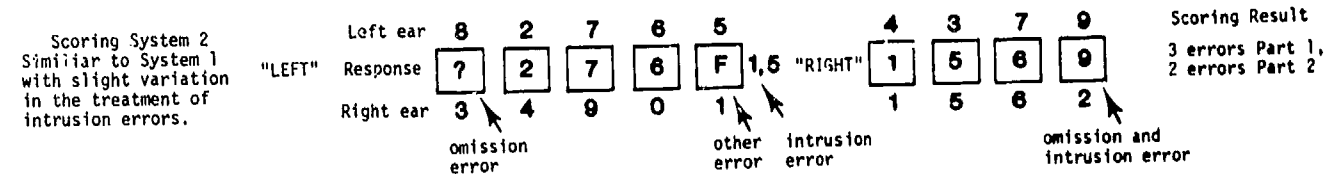
6. Report 3585, Cambridge, Mass:Bolt, Beranek, and Newman, Inc., November 1977.

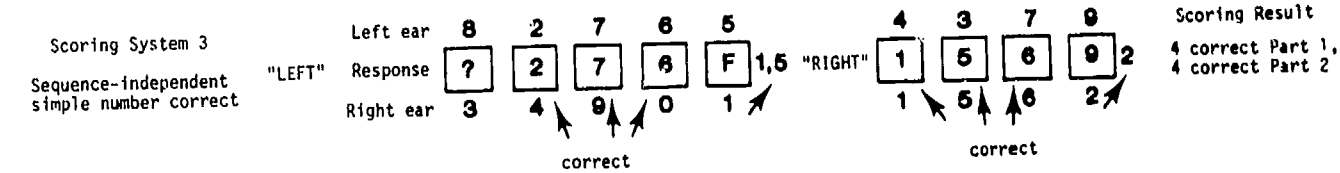
APPENDIX A

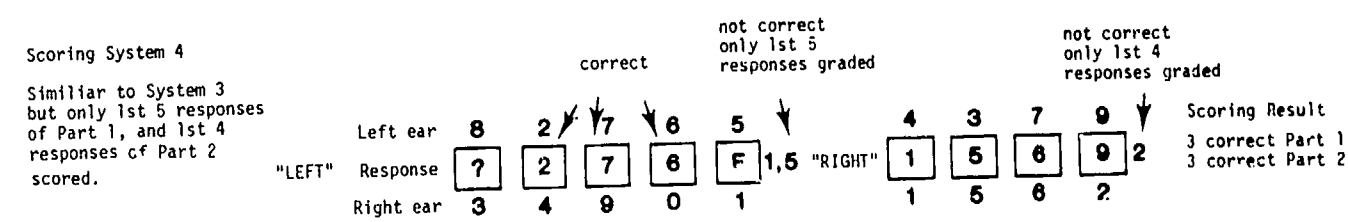
APPENDIX A

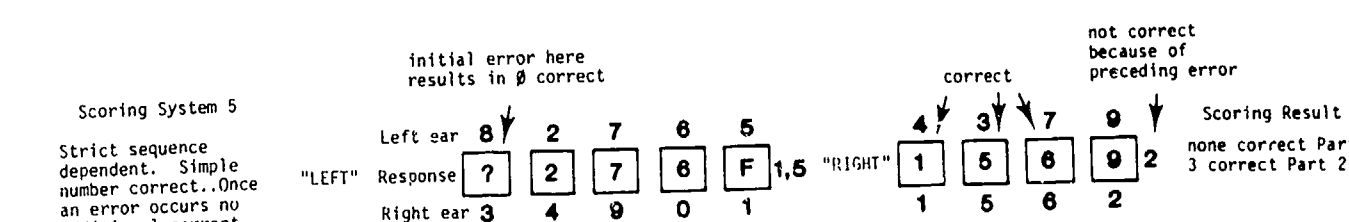
Results of the Application of Table I DLT Scoring Systems







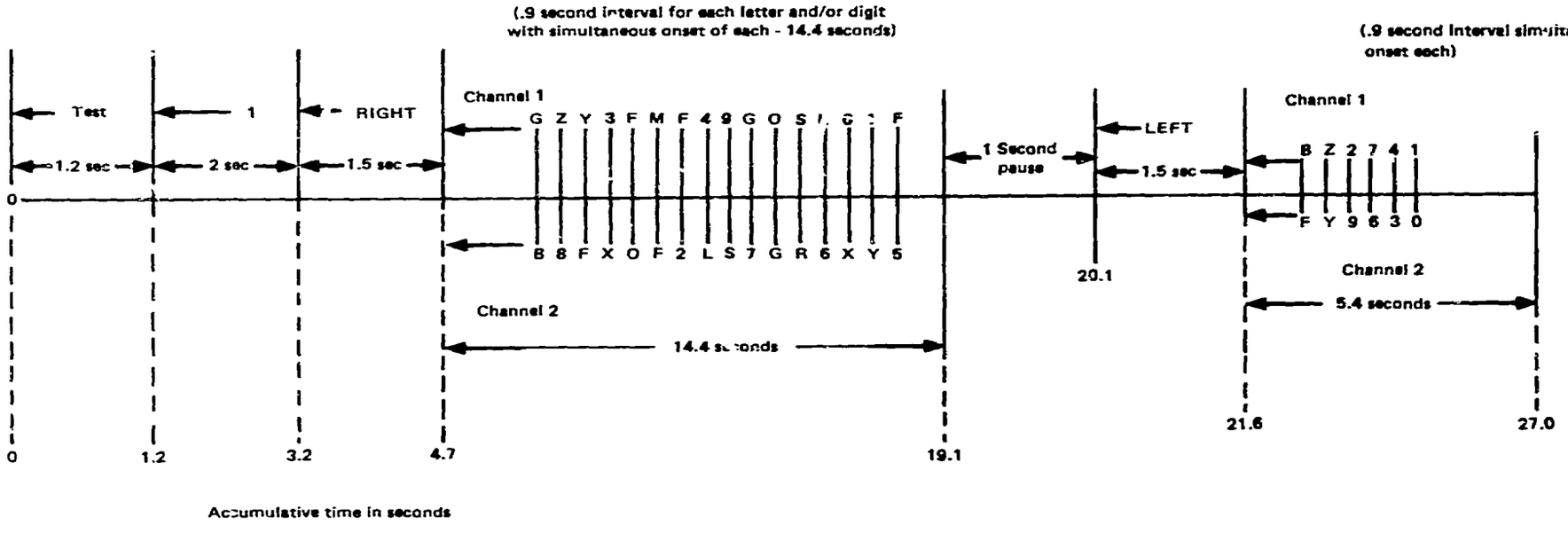




APPENDIX B

Appendix B

DLT Time Sequence Diagram



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Special Report 85-4	2. GOVT ACCESSION NO. AD-A159 920	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A comparison of dichotic listening task scoring methods		5. TYPE OF REPORT & PERIOD COVERED Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G. R. Griffin and J. D. Mosko		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Aerospace Medical Research Laboratory Naval Air Station, Pensacola, Florida 32508-5700		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Naval Medical Research and Development Command M0096001.1051 63706N
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command Naval Medical Command, National Capitol Region Bethesda, Maryland 20814		12. REPORT DATE March 1985 13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; dist ibution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Dichotic Listening
Selection
Prediction
Scoring Methods
Performance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In a recent evaluation of two dichotic listening tasks (DLTs) as predictors of performance in Naval Aviation Undergraduate Pilot Training, results based on one of five scoring methods were presented. The purpose of this report is to provide DLT

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (continued).

performance scores using all five scoring methods for comparative purposes to determine the scoring system most economical and efficient for automated scoring and, most sensitive to individual and mean differences.

Five scoring methods are described which vary primarily in their treatments of errors, and in their consideration of the importance of sequence effects. Five independent analyses of previously reported data were performed. Results were nearly invariant across all five methods; i.e., intercorrelations among scores across scoring techniques exceeded 0.90. Two of the scoring methods are recommended for purposes of standardizing future analyses of DLT performance, one because of its simplicity and ease of application and the other, because it may have greater sensitivity to differences in individual performance.